CLAIMS

What is claimed is:

1. A method for fast on-line automatic speaker/environment adaptation suitable for speech/speaker recognition in the presence of changing environmental conditions, the method comprising acts of:
   - performing front-end processing on an acoustic input signal, wherein the front-end processing generates MEL frequency cepstral features representative of the acoustic input signal;
   - performing recognition and adaptation by:
     - providing the MEL frequency cepstral features to a speech recognizer, wherein the speech recognizer utilizes the MEL frequency cepstral features and a current list of acoustic training models to determine at least one best hypothesis;
     - receiving, from the speech recognizer, at least one best hypothesis, associated acoustic training models, and associated probabilities;
     - computing a pre-adaptation acoustic score by recognizing an utterance using the associated acoustic training models;
     - choosing acoustic training models from the associated acoustic training models;
     - performing adaptation on the chosen associated acoustic training models;
     - computing a post-adaptation acoustic score by recognizing the utterance using the adapted acoustic training models;
     - comparing the pre-adaptation acoustic score with the post-adaptation acoustic score to check for improvement; modifying the current list of acoustic training models to include the adapted acoustic training models, if the acoustic score improved after performing adaptation;

and performing recognition and adaptation iteratively until the acoustic score ceases to improve;

- choosing the best hypothesis as recognized words once the acoustic score ceases to improve; and
- outputting the recognized words.

2. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 1, further comprising an act of receiving an acoustic input signal from an audio inputting device, the audio inputting device is selected from a group consisting of a microphone, a radio, a cellular wireless telephone, a telephone receiver, and an audio recording medium used to gather data in random environments and from non-standard speakers, the audio recording medium selected from a group consisting of an audio Compact Disk (CD), a cassette tape, a Digital Versatile Disk / Digital Video Disk (DVD), a video cassette, and a Long Play (LP) record.

3. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 2, wherein in the act of performing recognition and adaptation, the current list of acoustic training models available in the speech recognizer is comprised of a plurality of acoustic training models that are dependent on a speaker or on environmental conditions.

4. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 3, wherein, in the act of performing recognition and adaptation, the speech recognizer comprises a pattern matching act, a word generator act, and a sentence generator act.

5. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 4, wherein, in the act of performing recognition and adaptation, the speech recognizer inputs the MEL frequency cepstral features and the current list of acoustic training models into a pattern matching act, and the pattern matching act produces a set of units of sound.

6. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 5, wherein, in the act of performing recognition and adaptation, the pattern matching act of the speech recognizer comprises the acts of:

- generating a probability distribution function representing the inputted MEL frequency cepstral features;
- comparing the probability distribution function representing the inputted MEL frequency cepstral features with a plurality of probability distribution functions corresponding to all acoustic training models stored in the current list of acoustic training models; and
- selecting a set of units of sound that correspond to closer matches between the probability distribution function of the MEL frequency cepstral features and the probability distribution functions of all the models in the current list of acoustic training models.

7. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 6, wherein the MEL-frequency cepstral representation has acoustic landmarks of varying robustness, and where the pattern matching act locates the acoustic landmarks from the MEL-frequency cepstral representation, and embeds the acoustic landmarks into an acoustic network.

8. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 7, wherein the acoustic network includes segments, and wherein the pattern matching act further maps the segments in the acoustic network to units of sound hypotheses using a set of automatically determined acoustic parameters and acoustic training models in conjunction with pattern recognition algorithms.

9. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 8, wherein in the pattern matching act, a phoneme corresponds to a unit of sound and the pattern matching act outputs phoneme hypotheses.

10. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 9, further comprising an act of getting phonotactic models for the word generator act from a plurality of available phonotactic models, wherein the phonotactic models are independent from a speaker or from environmental conditions.

11. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 10, wherein the word generator act generates a set of word hypotheses by comparing the set of units of sound with the phonotactic models.

12. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 11, further comprising an act of getting language models for the sentence generator act from a plurality of available language models, wherein the language models are independent from a speaker or from environmental conditions.

13. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 12, wherein the sentence generator act generates a set of sentence hypotheses by comparing the set of word hypotheses and the language models.

14. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 13, wherein the speech recognizer outputs the set of sentence hypotheses produced by the sentence generator act, and a set of likelihood measures, wherein each likelihood measure is associated with a sentence hypothesis in the set of sentence hypotheses.

15. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 14, wherein the likelihood measure for each sentence hypothesis comprises a probability associated with a unit of sound, a set of probabilities associated with that unit of sound transitioning to several other units of sound, a probability associated with a word, and a set of probabilities associated with the word transitioning to several other words.

16. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 15, wherein the act of performing recognition and adaptation chooses a hypothesis with a highest likelihood measure to be a best hypothesis, and outputs at least one best hypothesis and its associated acoustic training models.

17. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 16, wherein, in the act of performing recognition and adaptation, the acoustic training models are stored in the current list of acoustic training models by grouping together the acoustic training models, representative of a unit of

sound, thus forming clusters of models, and wherein each cluster of models representative of a unit of sound has an outer layer.

18. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 17, wherein the act of performing recognition and adaptation further uses an Euclidean distance measure to select only the associated training models located on the outer layer of each cluster to be adapted, furthermore wherein the number of selected acoustic training models varies from utterance to utterance.

19. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 18, wherein the chosen acoustic training models associated with the best hypothesis have a set of mixture components, and wherein the act of adaptation estimates the distortion parameters for each chosen associated acoustic training model and for a chosen sub-set of mixture components from each chosen associated acoustic training model.

20. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 19, wherein each mixture component has a probability associated with it, and the chosen sub-set of mixture components is selected based on a fixed probability threshold value set a priori by a user, wherein only mixture components selected for adaptation are mixture components whose associated probability is at least equal to the chosen probability threshold value, and wherein the number of selected mixture components varies from utterance to utterance.

21. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 20, wherein, in the act of performing recognition and adaptation, the associated acoustic training models are adapted by incorporating distortion

parameters representative of current sound disturbances selected from a group consisting of changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

22. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 21, wherein the associated acoustic training models differ from acoustic training models representing the current sound disturbances, and wherein the distortion parameters consist of a bias mean value and a bias standard deviation value, representing differences between a mean and standard deviation of the associated acoustic training models, and a mean and standard deviation of the acoustic training models representing the current sound disturbances.

23. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 22, wherein the act of performing recognition and adaptation initializes the distortion parameters of the chosen associated acoustic training models to an initial bias mean value and an initial bias standard deviation value.

24. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 23, further comprising an act of computing an auxiliary function based on the initial bias mean value and the initial bias standard deviation value of the distortion parameters.

25. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 24, further comprising an act of iteratively performing Estimation Maximization of the auxiliary function over the chosen associated acoustic training models and mixture components, and wherein the Estimation

Maximization results in finding the distortion parameters that model most closely current environmental conditions, a present non-standard speaker acoustic model, or a present distorted sound acoustic model.

26. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 25, wherein the chosen associated acoustic training models is adapted by adding the distortion parameters, the distortion parameters consisting of a bias mean value and a bias standard deviation value, to the mean and standard deviation of the previously chosen associated acoustic training models, and wherein the chosen associated acoustic training models that have been adapted are labeled as adapted acoustic training models.

27. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 26, wherein in the act of performing recognition and adaptation, in the computing of an acoustic score, the acoustic training models include one of: associated acoustic training models or adapted acoustic training models; and the computing is performed by determining the best hypothesis using the acoustic training models and then combining a proper subset of the resulting associated probabilities from the best hypothesis, wherein the resulting associated probabilities from the best hypothesis used to determine the acoustic score comprise the probability associated with a unit of sound and a set of probabilities associated with that unit of sound transitioning to several other units of sound.

28. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 27, wherein in the act of outputting the recognized words, the outputting device is selected from a group consisting of a speaker-unit coupled with a computer system, a computer monitor, a electromagnetic wave representation of

an audio transmission, an audio Compact Disk (CD), a cassette tape, a Digital Versatile Disk/ Digital Video Disk (DVD), a video cassette, and a Long Play (LP) record.

29. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 1, wherein the act of performing recognition and adaptation further outputs the adapted acoustic training models that yielded the best hypothesis into a database of acoustic training models, wherein the database of acoustic training models will grow as new adapted acoustic training models generate new best hypothesis results for scenarios that include at least one of: changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

30. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 29, wherein the database of acoustic training models is tailored for non-standard speaker recognition, suitable for speech/speaker recognition applications comprising INS surveillance, national security surveillance, airport surveillance, automatic-speech telephone queries, air travel reservations, voice activated command and control systems, and automatic translation.

31. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 1, wherein, in the act of performing recognition and adaptation, the acoustic training models are stored in the current list of acoustic training models by grouping together the acoustic training models representative of a unit of sound, thus forming clusters of models, and wherein each cluster of models, representative of a unit of sound, has an outer layer.

32. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 31, wherein the act of performing recognition and adaptation further uses an Euclidean distance measure to select only the associated training models located on the outer layer of each cluster to be adapted, furthermore wherein the number of selected acoustic training models varies from utterance to utterance.

33. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 1, wherein the chosen acoustic training models associated with the best hypothesis have a set of mixture components, and wherein the act of adaptation estimates distortion parameters for each chosen associated acoustic training model and for a chosen sub-set of mixture components from each chosen associated acoustic training model.

34. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 33, wherein each mixture component has a probability associated with it, and the chosen sub-set of mixture components is selected based on a fixed probability threshold value set a priori by a user, wherein only mixture components selected for adaptation are mixture components whose associated probability is at least equal to the chosen probability threshold value, and wherein the number of selected mixture components varies from utterance to utterance.

35. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 1, wherein the associated acoustic training models differ from acoustic training models representing the current sound disturbances, and wherein distortion parameters consist of a bias mean value and a bias standard deviation value, representing differences between a mean and standard deviation of the

associated acoustic training models, and a mean and standard deviation of the acoustic training models representing the current sound disturbances.

36. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 35, wherein, in the act of performing recognition and adaptation, the associated acoustic training models are adapted by incorporating the distortion parameters representative of current sound disturbances selected from a group consisting of changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

37. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 36, wherein the act of performing recognition and adaptation initializes the distortion parameters of the chosen associated acoustic training models to an initial bias mean value and an initial bias standard deviation value.

38. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 37, further comprising an act of computing an auxiliary function based on the initial bias mean value and the initial bias standard deviation value of the distortion parameters.

39. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 38, further comprising an act of iteratively performing Estimation Maximization of the auxiliary function over the chosen associated acoustic training models and mixture components, and wherein the Estimation Maximization results in finding the distortion parameters that model most closely

current environmental conditions, a present non-standard speaker acoustic model, or a present distorted sound acoustic model.

40. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 36, wherein the chosen associated acoustic training models is adapted by adding the distortion parameters, the distortion parameters consisting of a bias mean value and a bias standard deviation value, to the mean and standard deviation of the previously chosen associated acoustic training models, and wherein the chosen associated acoustic training models that have been adapted are labeled as adapted acoustic training models.

41. A method for fast on-line automatic speaker/environment adaptation as set forth in claim 1, wherein in the act of performing recognition and adaptation, in the computing of an acoustic score, the acoustic training models include one of: associated acoustic training models or adapted acoustic training models; and the computing is performed by determining the best hypothesis using the acoustic training models and then combining a proper subset of the resulting associated probabilities from the best hypothesis, wherein the resulting associated probabilities from the best hypothesis used to determine the acoustic score comprise the probability associated with a unit of sound and a set of probabilities associated with that unit of sound transitioning to several other units of sound.

42. A system for fast on-line automatic speaker/environment adaptation suitable for speech/speaker recognition in the presence of changing environmental conditions, the system comprising:

a computer system including a processor, a memory coupled with the processor, an input coupled with the processor for receiving an acoustic input

signal, the computer system further comprising means, residing in its processor and memory for:

- performing front-end processing on the acoustic input signal, wherein the front-end processing generates MEL frequency cepstral features representative of the acoustic input signal;
- performing recognition and adaptation by:
    - providing the MEL frequency cepstral features to a speech recognizer, wherein the speech recognizer utilizes the MEL frequency cepstral features and a current list of acoustic training models to determine at least one best hypothesis;
    - receiving, from the speech recognizer, at least one best hypothesis, associated acoustic training models, and associated probabilities;
    - computing a pre-adaptation acoustic score by recognizing an utterance using the associated acoustic training models;
    - choosing acoustic training models from the associated acoustic training models;
    - performing adaptation on the chosen associated acoustic training models;
    - computing a post-adaptation acoustic score by recognizing the utterance using the adapted acoustic training models;
    - comparing the pre-adaptation acoustic score with the post-adaptation acoustic score to check for improvement; modifying the current list of acoustic training models to include the adapted acoustic training models, if the acoustic score improved after performing adaptation; and performing recognition and adaptation iteratively until the acoustic score ceases to improve;

- choosing the best hypothesis as recognized words once the acoustic score ceases to improve; and
- outputting the recognized words.

43. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 42, further comprising means for receiving the acoustic input signal from an audio inputting device, the audio inputting device is selected from a group consisting of a microphone, a radio, a cellular wireless telephone, a telephone receiver, and an audio recording medium used to gather data in random environments and from non-standard speakers, the audio recording medium selected from a group consisting of an audio Compact Disk (CD), a cassette tape, a Digital Versatile Disk / Digital Video Disk (DVD), a video cassette, and a Long Play (LP) record.

44. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 43, wherein in the means for performing recognition and adaptation, the current list of acoustic training models available in the speech recognizer is comprised of a plurality of acoustic training models that are dependent on a speaker or on environmental conditions.

45. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 44, wherein, in the means for performing recognition and adaptation, the speech recognizer comprises means for:

- pattern matching;
- word generation; and
- sentence generation.

46. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 45, wherein, in the means for performing recognition and adaptation, the speech recognizer inputs the MEL frequency cepstral features and the current list of acoustic training models into a pattern matching act, and the pattern matching act produces a set of units of sound.

47. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 46, wherein, in the means for performing recognition and adaptation, the means for pattern matching of the speech recognizer comprises the means for:

- generating a probability distribution function representing the inputted MEL frequency cepstral features;
- comparing the probability distribution function representing the inputted MEL frequency cepstral features with a plurality of probability distribution functions corresponding to all acoustic training models stored in the current list of acoustic training models; and
- selecting a set of units of sound that correspond to closer matches between the probability distribution function of the MEL frequency cepstral features and the probability distribution functions of all the models in the current list of acoustic training models.

48. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 47, wherein the MEL-frequency cepstral representation has acoustic landmarks of varying robustness, and where the means for pattern matching locates the acoustic landmarks from the MEL-frequency cepstral representation, and embeds the acoustic landmarks into an acoustic network.

49. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 48, wherein the acoustic network includes segments, and wherein the means for pattern matching further maps the segments in the acoustic network to units of sound hypotheses using a set of automatically determined acoustic parameters and acoustic training models in conjunction with pattern recognition algorithms.

50. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 49, wherein in the means for pattern matching, a phoneme corresponds to a unit of sound and the means for pattern matching outputs phoneme hypotheses.

51. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 50, further comprising means for getting phonotactic models for the means for word generation from a plurality of available phonotactic models, wherein the phonotactic models are independent from a speaker or from environmental conditions.

52. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 51, wherein the means for word generation generates a set of word hypotheses by comparing the set of units of sound with the phonotactic models.

53. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 52, further comprising means for getting language models for the means for sentence generation from a plurality of available language models, wherein the language models are independent from a speaker or from environmental conditions.

54. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 53, wherein the means for sentence generation generates a set of sentence hypotheses by comparing the set of word hypotheses and the language models.

55. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 54, wherein the speech recognizer outputs the set of sentence hypotheses produced by the means for sentence generation, and a set of likelihood measures, wherein each likelihood measure is associated with a sentence hypothesis in the set of sentence hypotheses.

56. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 55, wherein the likelihood measure for each sentence hypothesis comprises a probability associated with a unit of sound, a set of probabilities associated with that unit of sound transitioning to several other units of sound, a probability associated with a word, and a set of probabilities associated with the word transitioning to several other words.

57. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 56, wherein the means for performing recognition and adaptation chooses a hypothesis with a highest likelihood measure to be a best hypothesis, and outputs at least one best hypothesis and its associated acoustic training models.

58. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 57, wherein, in the means for performing recognition and adaptation, the acoustic training models are stored in the current list of acoustic training models by grouping together the acoustic training models representative of a unit of

sound, thus forming clusters of models, and wherein each cluster of models representative of a unit of sound has an outer layer.

59. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 58, wherein the means for performing recognition and adaptation further uses an Euclidean distance measure to select only the associated training models located on the outer layer of each cluster to be adapted, furthermore wherein the number of selected acoustic training models varies from utterance to utterance.

60. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 59, wherein the chosen acoustic training models associated with the best hypothesis have a set of mixture components, and wherein the means for adaptation estimates the distortion parameters for each chosen associated acoustic training model and for a chosen sub-set of mixture components from each chosen associated acoustic training model.

61. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 60, wherein each mixture component has a probability associated with it, and the chosen sub-set of mixture components is selected based on a fixed probability threshold value set a priori by a user, wherein only mixture components selected for adaptation are mixture components whose associated probability is at least equal to the chosen probability threshold value, and wherein the number of selected mixture components varies from utterance to utterance.

62. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 61, wherein, in the means for performing recognition and adaptation, the associated acoustic training models are adapted by incorporating distortion

parameters representative of current sound disturbances selected from a group consisting of changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

63. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 62, wherein the associated acoustic training models differ from acoustic training models representing the current sound disturbances, and wherein the distortion parameters consist of a bias mean value and a bias standard deviation value, representing differences between a mean and standard deviation of the associated acoustic training models, and a mean and standard deviation of the acoustic training models representing the current sound disturbances.

64. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 63, wherein the means for performing recognition and adaptation initializes the distortion parameters of the chosen associated acoustic training models to an initial bias mean value and an initial bias standard deviation value.

65. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 64, further comprising means for computing an auxiliary function based on the initial bias mean value and the initial bias standard deviation value of the distortion parameters.

66. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 65, further comprising means for iteratively performing Estimation Maximization of the auxiliary function over the chosen associated acoustic training models and mixture components, and wherein the Estimation

Maximization results in finding the distortion parameters that model most closely current environmental conditions, a present non-standard speaker acoustic model, or a present distorted sound acoustic model.

67. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 66, wherein the chosen subset of associated acoustic training models is adapted by adding the distortion parameters, the distortion parameters consisting of a bias mean value and a bias standard deviation value, to the mean and standard deviation of the previously chosen associated acoustic training models, and wherein the chosen associated acoustic training models that have been adapted are labeled as adapted acoustic training models.

68. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 67, wherein in the means for performing recognition and adaptation, in the computing of an acoustic score, the acoustic training models include one of: associated acoustic training models or adapted acoustic training models; and the computing is performed by determining the best hypothesis using the acoustic training models and then combining a proper subset of the resulting associated probabilities from the best hypothesis, wherein the resulting associated probabilities from the best hypothesis used to determine the acoustic score comprise the probability associated with a unit of sound and a set of probabilities associated with that unit of sound transitioning to several other units of sound.

69. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 68, wherein in the means for outputting the recognized words, the outputting device is selected from a group consisting of a speaker-unit coupled with a computer system, a computer monitor, a electromagnetic wave

representation of an audio transmission, an audio Compact Disk (CD), a cassette tape, a Digital Versatile Disk/ Digital Video Disk (DVD), a video cassette, and a Long Play (LP) record.

70. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 42, wherein the means for performing recognition and adaptation further outputs the adapted acoustic training models that yielded the best hypothesis into a database of acoustic training models, wherein the database of acoustic training models will grow as new adapted acoustic training models generate new best hypothesis results for scenarios that include at least one of: changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

71. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 70, wherein the database of acoustic training models is tailored for non-standard speaker recognition, suitable for speech/speaker recognition applications comprising INS surveillance, national security surveillance, airport surveillance, automatic-speech telephone queries, air travel reservations, voice activated command and control systems, and automatic translation.

72. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 42, wherein, in the means for performing recognition and adaptation, the acoustic training models are stored in the current list of acoustic training models by grouping together the acoustic training models representative of a unit of sound, thus forming clusters of models, and wherein each cluster of models representative of a unit of sound has an outer layer.

73. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 72, wherein the means for performing recognition and adaptation further uses an Euclidean distance measure to select only the associated training models located on the outer layer of each cluster to be adapted, furthermore wherein the number of selected acoustic training models varies from utterance to utterance.

74. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 42, wherein the chosen acoustic training models associated with the best hypothesis have a set of mixture components, and wherein the means for adaptation estimates distortion parameters for each chosen associated acoustic training model and for a chosen sub-set of mixture components from each chosen associated acoustic training model.

75. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 74, wherein each mixture component has a probability associated with it, and the chosen sub-set of mixture components is selected based on a fixed probability threshold value set a priori by a user, wherein only mixture components selected for adaptation are mixture components whose associated probability is at least equal to the chosen probability threshold value, and wherein the number of selected mixture components varies from utterance to utterance.

76. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 42, wherein the associated acoustic training models differ from acoustic training models representing the current sound disturbances, and wherein distortion parameters consist of a bias mean value and a bias standard deviation value, representing differences between a mean and standard deviation of the

associated acoustic training models, and a mean and standard deviation of the acoustic training models representing the current sound disturbances.

77. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 76, wherein, in the means for performing recognition and adaptation, the associated acoustic training models are adapted by incorporating the distortion parameters representative of current sound disturbances selected from a group consisting of changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

78. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 77, wherein the means for performing recognition and adaptation initializes the distortion parameters of the chosen associated acoustic training models to an initial bias mean value and an initial bias standard deviation value.

79. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 78, further comprising means for computing an auxiliary function based on the initial bias mean value and the initial bias standard deviation value of the distortion parameters.

80. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 79, further comprising means for iteratively performing Estimation Maximization of the auxiliary function over the chosen subsets of associated acoustic training models and mixture components, and wherein the Estimation Maximization results in finding the distortion parameters that model most closely

current environmental conditions, a present non-standard speaker acoustic model, or a present distorted sound acoustic model.

81. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 77, wherein the chosen subset of associated acoustic training models is adapted by adding the distortion parameters, the distortion parameters consisting of a bias mean value and a bias standard deviation value, to the mean and standard deviation of the previously chosen associated acoustic training models, and wherein the chosen associated acoustic training models that have been adapted are labeled as adapted acoustic training models.

82. A system for fast on-line automatic speaker/environment adaptation as set forth in claim 42, wherein in the means for performing recognition and adaptation, in the computing of an acoustic score, the acoustic training models include one of: associated acoustic training models or adapted acoustic training models; and the computing is performed by determining the best hypothesis using the acoustic training models and then combining a proper subset of the resulting associated probabilities from the best hypothesis, wherein the resulting associated probabilities from the best hypothesis used to determine the acoustic score comprise the probability associated with a unit of sound and a set of probabilities associated with that unit of sound transitioning to several other units of sound.

83. A computer program product for fast on-line automatic speaker/environment adaptation suitable for speech/speaker recognition in the presence of changing environmental conditions, the computer program product comprising means, stored on a computer readable medium for:

- receiving an acoustic input signal;

- performing front-end processing on the acoustic input signal, wherein the front-end processing generates MEL frequency cepstral features representative of the acoustic input signal;
- performing recognition and adaptation by:
    - providing the MEL frequency cepstral features to a speech recognizer, wherein the speech recognizer utilizes the MEL frequency cepstral features and a current list of acoustic training models to determine at least one best hypothesis;
    - receiving, from the speech recognizer, at least one best hypothesis, associated acoustic training models, and associated probabilities;
    - computing a pre-adaptation acoustic score by recognizing the utterance using the associated acoustic training models;
    - choosing acoustic training models from the associated acoustic training models;
    - performing adaptation on the chosen associated acoustic training models;
    - computing a post-adaptation acoustic score by recognizing the utterance using the adapted acoustic training models;
    - comparing the pre-adaptation acoustic score with the post-adaptation acoustic score to check for improvement; modifying the current list of acoustic training models to include the adapted acoustic training models, if the acoustic score improved after performing adaptation; and performing recognition and adaptation iteratively until the acoustic score ceases to improve;
- choosing the best hypothesis as recognized words once the acoustic score ceases to improve; and
- outputting the recognized words.

84. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 83, wherein the means for receiving an acoustic input signal from an audio inputting device, the audio inputting device is selected from a group consisting of a microphone, a radio, a cellular wireless telephone, a telephone receiver, and an audio recording medium used to gather data in random environments and from non-standard speakers, the audio recording medium selected from a group consisting of an audio Compact Disk (CD), a cassette tape, a Digital Versatile Disk / Digital Video Disk (DVD), a video cassette, and a Long Play (LP) record.

85. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 84, wherein in the means for performing recognition and adaptation, the current list of acoustic training models available in the speech recognizer is comprised of a plurality of acoustic training models that are dependent on a speaker or on environmental conditions.

86. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 85, wherein, in the means for performing recognition and adaptation, the speech recognizer comprises means for:

- pattern matching;
- word generation; and
- sentence generation.

87. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 86, wherein, in the means for performing recognition and adaptation, the speech recognizer inputs the MEL frequency

cepstral features and the current list of acoustic training models into a pattern matching act, and the pattern matching act produces a set of units of sound.

88. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 87, wherein, in the means for performing recognition and adaptation, the means for pattern matching of the speech recognizer comprises the means for:

- generating a probability distribution function representing the inputted MEL frequency cepstral features;
- comparing the probability distribution function representing the inputted MEL frequency cepstral features with a plurality of probability distribution functions corresponding to all acoustic training models stored in the current list of acoustic training models; and
- selecting a set of units of sound that correspond to closer matches between the probability distribution function of the MEL frequency cepstral features and the probability distribution functions of all the models in the current list of acoustic training models.

89. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 88, wherein the MEL-frequency cepstral representation has acoustic landmarks of varying robustness, and where the means for pattern matching locates the acoustic landmarks from the MEL-frequency cepstral representation, and embeds the acoustic landmarks into an acoustic network.

90. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 89, wherein the acoustic network includes

segments, and wherein the means for pattern matching further maps the segments in the acoustic network to units of sound hypotheses using a set of automatically determined acoustic parameters and acoustic training models in conjunction with pattern recognition algorithms.

91. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 90, wherein in the means for pattern matching, a phoneme corresponds to a unit of sound and the means for pattern matching outputs phoneme hypotheses.

92. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 91, further comprising means for getting phonotactic models for the means for word generation from a plurality of available phonotactic models, wherein the phonotactic models are independent from a speaker or from environmental conditions.

93. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 92, wherein the means for word generation generates a set of word hypotheses by comparing the set of units of sound with the phonotactic models.

94. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 93, further comprising means for getting language models for the means for sentence generation from a plurality of available language models, wherein the language models are independent from a speaker or from environmental conditions.

95. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 94, wherein the means for sentence generation generates a set of sentence hypotheses by comparing the set of word hypotheses and the language models.

96. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 95, wherein the speech recognizer outputs the set of sentence hypotheses produced by the means for sentence generation, and a set of likelihood measures, wherein each likelihood measure is associated with a sentence hypothesis in the set of sentence hypotheses.

97. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 96, wherein the likelihood measure for each sentence hypothesis comprises a probability associated with a unit of sound, a set of probabilities associated with that unit of sound transitioning to several other units of sound, a probability associated with a word, and a set of probabilities associated with the word transitioning to several other words.

98. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 97, wherein the means for performing recognition and adaptation chooses a hypothesis with a highest likelihood measure to be a best hypothesis, and outputs at least one best hypothesis and its associated acoustic training models.

99. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 98, wherein, in the means for performing recognition and adaptation, the acoustic training models are stored in the current

list of acoustic training models by grouping together the acoustic training models representative of a unit of sound, thus forming clusters of models, and wherein each cluster of models representative of a unit of sound has an outer layer.

100. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 99, wherein the means for performing recognition and adaptation further uses an Euclidean distance measure to select only the associated training models located on the outer layer of each cluster to be adapted, furthermore wherein the number of selected acoustic training models varies from utterance to utterance.

101. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 100, wherein the chosen acoustic training models associated with the best hypothesis have a set of mixture components, and wherein the means for adaptation estimates distortion parameters for each chosen associated acoustic training model and for a chosen sub-set of mixture components from each chosen associated acoustic training model.

102. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 101, wherein each mixture component has a probability associated with it, and the chosen sub-set of mixture components is selected based on a fixed probability threshold value set a priori by a user, wherein only mixture components selected for adaptation are mixture components whose associated probability is at least equal to the chosen fixed probability threshold value, and wherein the number of selected mixture components varies from utterance to utterance.

103. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 102, wherein, in the means for performing recognition and adaptation, the associated acoustic training models are adapted by incorporating distortion parameters representative of current sound disturbances selected from a group consisting of changing environmental conditions, deviation of a speaker from the standard language, and deviation of a sound from the standard sound characteristics.

104. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 103, wherein the associated acoustic training models differ from acoustic training models representing the current sound disturbances, and wherein the distortion parameters consist of a bias mean value and a bias standard deviation value, representing differences between a mean and standard deviation of the associated acoustic training models, and a mean and standard deviation of the acoustic training models representing the current sound disturbances.

105. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 104, wherein the means for performing recognition and adaptation initializes the distortion parameters of the chosen associated acoustic training models to an initial bias mean value and an initial bias standard deviation value.

106. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 105, further comprising means for computing an auxiliary function based on the initial bias mean value and the initial bias standard deviation value of the distortion parameters.

107. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 106, further comprising means for iteratively performing Estimation Maximization of the auxiliary function over the chosen subsets of associated acoustic training models and mixture components, and wherein the Estimation Maximization results in finding the distortion parameters that model most closely current environmental conditions, a present non-standard speaker acoustic model, or a present distorted sound acoustic model.

108. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 107, wherein the chosen subset of associated acoustic training models is adapted by adding the distortion parameters, the distortion parameters consisting of a bias mean value and a bias standard deviation value, to the mean and standard deviation of the previously chosen associated acoustic training models, and wherein the chosen associated acoustic training models that have been adapted are labeled as adapted acoustic training models.

109. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 108, wherein in the means for performing recognition and adaptation, in the computing of an acoustic score, the acoustic training models include one of: associated acoustic training models or adapted acoustic training models; and the computing is performed by determining the best hypothesis using the acoustic training models and then combining a proper subset of the resulting associated probabilities from the best hypothesis, wherein the resulting associated probabilities from the best hypothesis used to determine the acoustic score comprise the probability associated with a unit of sound and a set

of probabilities associated with that unit of sound transitioning to several other units of sound.

110. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 109, wherein, in the means for outputting the recognized words, the outputting device is selected from a group consisting of a speaker-unit coupled with a computer system, a computer monitor, a electromagnetic wave representation of an audio transmission, an audio Compact Disk (CD), a cassette tape, a Digital Versatile Disk/ Digital Video Disk (DVD), a video cassette, and a Long Play (LP) record.

111. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 83, wherein the means for performing recognition and adaptation further outputs the adapted acoustic training models that yielded the best hypothesis into a database of acoustic training models, wherein the database of acoustic training models grows as new adapted acoustic training models generate new best hypothesis results for scenarios that include at least one of: changing environmental conditions, deviation of a speaker from standard language, and deviation of a sound from standard sound characteristics.

112. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 111, wherein the database of acoustic training models is tailored for non-standard speaker recognition, suitable for speech/speaker recognition applications comprising INS surveillance, national security surveillance, airport surveillance, automatic-speech telephone queries, air travel reservations, voice activated command and control systems, and automatic translation.

113. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 83, wherein, in the means for performing recognition and adaptation, the acoustic training models are stored in the current list of acoustic training models by grouping together the acoustic training models representative of a unit of sound, thus forming clusters of models, and wherein each cluster of models representative of a unit of sound has an outer layer.

114. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 113, wherein the means for performing recognition and adaptation further uses an Euclidean distance measure to select only the associated training models located on the outer layer of each cluster to be adapted, furthermore wherein a number of selected acoustic training models varies from utterance to utterance.

115. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 83, wherein the chosen acoustic training models associated with the best hypothesis have a set of mixture components, and wherein the means for adaptation estimates distortion parameters for each chosen associated acoustic training model and for a chosen sub-set of mixture components from each chosen associated acoustic training model.

116. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 115, wherein each mixture component has a probability associated with it, and the chosen sub-set of mixture components is selected based on a fixed probability threshold value set a priori by a user, wherein only mixture components selected for adaptation are mixture components

whose associated probability is at least equal to the chosen probability threshold value, and wherein the number of selected mixture components varies from utterance to utterance.

117. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 83, wherein the associated acoustic training models differ from acoustic training models representing current sound disturbances, and wherein distortion parameters consist of a bias mean value and a bias standard deviation value, representing differences between a mean and standard deviation of the associated acoustic training models, and a mean and standard deviation of the acoustic training models representing the current sound disturbances.

118. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 117, wherein, in the means for performing recognition and adaptation, the associated acoustic training models are adapted by incorporating the distortion parameters representative of current sound disturbances selected from a group consisting of changing environmental conditions, deviation of a speaker from standard language, and deviation of a sound from standard sound characteristics.

119. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 118, wherein the means for performing recognition and adaptation initializes the distortion parameters of the chosen associated acoustic training models to an initial bias mean value and an initial bias standard deviation value.

120. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 119, further comprising means for computing an auxiliary function based on the initial bias mean value and the initial bias standard deviation value of the distortion parameters.

121. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 120, further comprising means for iteratively performing Estimation Maximization of the auxiliary function over the chosen subsets of associated acoustic training models and mixture components, and wherein the Estimation Maximization results in finding the distortion parameters that model most closely current environmental conditions, a present non-standard speaker acoustic model, or a present distorted sound acoustic model.

122. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 118, wherein the chosen subset of associated acoustic training models is adapted by adding the distortion parameters, the distortion parameters consisting of a bias mean value and a bias standard deviation value, to the mean and standard deviation of the previously chosen associated acoustic training models, and wherein the chosen associated acoustic training models that have been adapted are labeled as adapted acoustic training models.

123. A computer program product for fast on-line automatic speaker/environment adaptation as set forth in claim 83, wherein, in the means for performing recognition and adaptation, in the computing of an acoustic score, the acoustic training models include one of: associated acoustic training models or adapted acoustic training models; and the computing is performed by determining the best

hypothesis using the acoustic training models and then combining a proper subset of resulting associated probabilities from the best hypothesis, wherein the resulting associated probabilities from the best hypothesis used to determine the acoustic score comprise the probability associated with a unit of sound and a set of probabilities associated with that unit of sound transitioning to several other units of sound.